Econometrica Supplementary Material

SUPPLEMENT TO "INFERENCE ON COUNTERFACTUAL DISTRIBUTIONS"
(*Econometrica*, Vol. 81, No. 6, November 2013, 2205–2268)

BY VICTOR CHERNOZHUKOV, IVÁN FERNÁNDEZ-VAL, AND BLAISE MELLY

This supplemental material contains four appendices. In Appendix SA, we show the validity of bootstrap confidence bands for general counterfactual functionals. In Appendix SB, we compare the performance of quantile and distribution regression as estimators of the conditional and counterfactual distribution functions in Monte Carlo simulations. We calibrate our data generating process to fit several characteristics of the Current Population Survey data used in Section 6 of the paper. We find that quantile regression works better than distribution regression when the distribution of the dependent variable is continuous, but it performs worse in the presence of a small and realistic amount of rounding in the data. Appendices SC and SD complement the empirical results presented in Section 6. Appendix SC contains descriptive statistics, additional results for men, and all the results for women. Appendix SD gives more details about the variance decomposition of the composition effect into between-group and within-group components.

KEYWORDS: Counterfactual distribution, quantile regression, distribution regression, bootstrap, minimum wage, unions.

APPENDIX SA: VALIDITY OF BOOTSTRAP CONFIDENCE BANDS AND UNIFORM CONSISTENCY OF BOOTSTRAP VARIANCE ESTIMATORS

WE CONSIDER A GENERAL COUNTERFACTUAL FUNCTIONAL

$$\Delta(w) = \phi\big(F_{Y\langle j|k\rangle} : (j,k) \in \mathcal{JK}\big)(w)$$

and its plug-in estimator

$$\widehat{\Delta}(w) = \phi\big(\widehat{F}_{Y\langle j|k\rangle} : (j,k) \in \mathcal{JK}\big)(w)$$

for $w \in \mathcal{W}$. In Sections 4 and 5 of the main paper (hereafter CFM), we established conditions for the functional central limit theorem,

$$\widehat{Z} := \sqrt{n}(\widehat{\Delta} - \Delta) \rightsquigarrow Z \quad \text{in} \quad \ell^\infty(\mathcal{W}),$$

where Z is a zero-mean Gaussian process with a.s. continuous sample paths and pointwise variance function $\Sigma(w) = E[Z^2(w)]$. Let

$$\widehat{\Delta}^*(w) = \phi\big(\widehat{F}^*_{Y\langle j|k\rangle} : (j,k) \in \mathcal{JK}\big)(w)$$

be the bootstrap version of $\widehat{\Delta}(w)$. We also gave conditions for the bootstrap functional central limit theorem:

$$\widehat{Z}^* := \sqrt{n}\big(\widehat{\Delta}^* - \widehat{\Delta}\big) \rightsquigarrow_{\mathbb{P}} Z \quad \text{in} \quad \ell^\infty(\mathcal{W}).$$

 DOI: 10.3982/ECTA10582

Let $\widehat{t}^*_{1-\alpha}$ be the $1-\alpha$ quantile of the bootstrap version of the Kolmogorov–Smirnov maximal t-statistic,

$$\widehat{t}^* = \sup_{w\in\mathcal{W}} \widehat{\Sigma}^*(w)^{-1/2}\big|\widehat{Z}^*(w)\big|,$$

conditional on the data. As a robust estimator of $\Sigma(w)$, we consider the rescaled squared bootstrap quantile spread,

$$\text{(SA.1)}\qquad \widehat{\Sigma}^*(w) := \big[\widehat{Z}^*_{\tau_1}(w) - \widehat{Z}^*_{\tau_2}(w)\big]^2/[Z_{\tau_1} - Z_{\tau_2}]^2,$$

where $\widehat{Z}^*_\tau(w)$ is the τ quantile of $\widehat{Z}^*(w)$ conditional on the data, Z_τ is the τ quantile of the $N(0,1)$, and $0<\tau_2<\tau_1<1$. The bootstrap versions of the end-point functions of the $1-\alpha$ confidence band are

$$\widehat{\Delta}^{\pm*}(w) = \widehat{\Delta}(w) \pm \widehat{t}^*_{1-\alpha}\widehat{\Sigma}^*(w)^{1/2}/\sqrt{n}.$$

The following lemma establishes that the bootstrap confidence band $[\widehat{\Delta}^{-*}(w), \widehat{\Delta}^{+*}(w)]$ covers $\Delta(w)$ uniformly over $w\in\mathcal{W}$ with asymptotic probability $1-\alpha$. It is based on Theorems 1 and 2, and Lemma 1 in Chernozhukov and Fernández-Val (2005), adapted to our setting.

LEMMA SA.1—Consistency of Bootstrap Confidence Bands: *Suppose that* $\widehat{Z} \rightsquigarrow Z$ *and* $\widehat{Z}^* \rightsquigarrow_{\mathbb{P}} Z$ *in* $\ell^\infty(\mathcal{W})$, *where* Z *is a tight zero-mean Gaussian process with continuous paths in* $\mathcal{W}$ *and variance function* $\Sigma(w)$ *that is bounded away from zero and from above uniformly over* $w\in\mathcal{W}$. *Then* (a) *the variance estimator specified in* (SA.1) *is uniformly consistent*,

$$\sup_{w\in\mathcal{W}}\big|\widehat{\Sigma}^*(w) - \Sigma(w)\big| \to_{\mathbb{P}} 0,$$

and (b)

$$\lim_{n\to\infty}\mathbb{P}\big\{\Delta(w)\in\big[\widehat{\Delta}^{-*}(w), \widehat{\Delta}^{+*}(w)\big]\textit{ for all } w\in\mathcal{W}\big\} = 1-\alpha.$$

To state the proof formally, we follow the notation and definitions in van der Vaart and Wellner (1996). Let D_n denote the data vector and let M_n be the vector of random variables used to generate bootstrap draws or simulation draws given D_n (this may depend on the particular resampling or simulation method). Consider the random element $\mathbb{Z}^*_n = \mathbb{Z}_n(D_n, M_n)$ in a normed space $\mathbb{D}$. We say that the bootstrap law of $\mathbb{Z}^*_n$ consistently estimates the law of some tight random element $\mathbb{Z}$ and we write $\mathbb{Z}^*_n \rightsquigarrow_{\mathbb{P}} \mathbb{Z}$ in $\mathbb{D}$ if

$$\text{(SA.2)}\qquad d_{\mathrm{BL}}\big(\mathbb{Z}^*_n, \mathbb{Z}\big) := \sup_{h\in\mathrm{BL}_1(\mathbb{D})}\big|E_{M_n}h\big(\mathbb{Z}^*_n\big) - Eh(\mathbb{Z})\big| \to_{\mathbb{P}} 0,$$

where $\mathrm{BL}_1(\mathbb{D})$ denotes the space of functions that map $\mathbb{D}$ to the real line with Lipschitz norm at most 1 and where E_{M_n} denotes the conditional expectation with respect to M_n given the data D_n.

PROOF OF LEMMA SA.1:

STEP 1: Here we establish claim (a). Note that for each $w \in \mathcal{W}$, $Z(w) \sim N(0, \Sigma(w))$, so that the τ quantile of $Z(w)$ is $Z_\tau(w) = \Sigma(w)^{1/2} Z_\tau$, and $\Sigma(w) = [Z_{\tau_1}(w) - Z_{\tau_2}(w)]^2/[Z_{\tau_1} - Z_{\tau_2}]^2$. Let

$$F(\nu, w) := P\big[Z(w) \leq \nu\big] = \Phi\big(\Sigma^{-1/2}(w)\nu\big),$$

where Φ is the c.d.f. of the $N(0,1)$. Note that $w \mapsto \Sigma(w)^{1/2}$ is continuous in $w \in \mathcal{W}$ by Gaussianity of Z and a.s. continuity of its sample paths with respect to $w \in \mathcal{W}$. Moreover, $w \mapsto \Sigma^{-1/2}(w)$ is uniformly continuous on $\mathcal{W}$ due to the continuity of $w \mapsto \Sigma^{1/2}(w)$ and $\Sigma^{1/2}(w)$ being bounded away from zero uniformly over $\mathcal{W}$. This implies that $(\nu, w) \mapsto F(\nu, w)$ is uniformly continuous on $(\nu, w) \in \mathbb{R} \times \mathcal{W}$ and that $Z_\tau(w) = \inf_{\nu \in \mathbb{R}}\{F(\nu, w) \geq \tau\}$ is uniformly continuous in $(\tau, w) \in \mathcal{T} \times \mathcal{W}$, where $\mathcal{T}$ is a compact subinterval of $(0,1)$. In the paragraph below, we deduce that $\widehat{F}(\nu, w) := E_{M_n} 1\{\widehat{Z}^*(w) \leq \nu\}$ obeys

$$\widehat{F}(\nu, w) \to_{\mathbb{P}} F(\nu, w) \quad \text{uniformly in } (\nu, w) \in \mathbb{R} \times \mathcal{W}. \tag{SA.3}$$

By Lemma 1 in Chernozhukov and Fernández-Val (2005), these properties imply that

$$\widehat{Z}^*_\tau(w) \to_{\mathbb{P}} Z_\tau(w) \quad \text{uniformly in } (\tau, w) \in \mathcal{T} \times \mathcal{W}. \tag{SA.4}$$

Therefore, uniformly in $w \in \mathcal{W}$, by the continuous mapping theorem,

$$\widehat{\Sigma}^*(w) = \frac{[\widehat{Z}^*_{\tau_1}(w) - \widehat{Z}^*_{\tau_2}(w)]^2}{[Z_{\tau_1} - Z_{\tau_2}]^2} \to_{\mathbb{P}} \frac{[Z_{\tau_1}(w) - Z_{\tau_2}(w)]^2}{[Z_{\tau_1} - Z_{\tau_2}]^2} = \Sigma(w).$$

The argument to show (SA.3) is based in part on the proof of Theorem 2 in Chernozhukov and Fernández-Val (2005). Let $u \mapsto K_{\delta,\nu}(u) := 1(u \leq \nu) + 1(\nu < u \leq \nu + \delta)(\nu + \delta - u)/\delta$ denote a smoothed approximation to the indicator function $u \mapsto 1(u \leq \nu)$, with $\delta > 0$ denoting the smoothing parameter. Then the collection of maps that map $z \in \ell^\infty(\mathcal{W})$ to $\mathbb{R}$, defined as $\mathcal{G}_\delta = \{z \mapsto K_{\delta,\nu}(z(w)) : \nu \in \mathbb{R}, w \in \mathcal{W}\}$, is a subset of $\delta^{-1}\,\mathrm{BL}_1(\ell^\infty(\mathcal{W}))$. By $\widehat{Z}^* \rightsquigarrow_{\mathbb{P}} Z$ in $\ell^\infty(\mathcal{W})$, $\sup_{g \in \mathcal{G}_\delta} |E_{M_n} g(\widehat{Z}^*) - Eg(Z)| \leq \delta^{-1} d_{\mathrm{BL}}(\widehat{Z}^*, Z) \to_{\mathbb{P}} 0$. Therefore, uniformly in $w \in \mathcal{W}$ and $\nu \in \mathbb{R}$,

$$\widehat{F}(\nu, w) = E_{M_n} 1\big\{\widehat{Z}^*(w) \leq \nu\big\} \leq E_{M_n} K_{\delta,\nu}\big[\widehat{Z}^*(w)\big] \to_{\mathbb{P}} EK_{\delta,\nu}\big[Z(w)\big]. \tag{SA.5}$$

We have that $F(\nu, w) \leq EK_{\delta,\nu}[Z(w)] \leq F(\nu + \delta, w)$. By uniform continuity of $F(\cdot,\cdot)$, as $\delta \searrow 0$, $EK_{\delta,\nu}[Z(w)] \to F(\nu, w)$ uniformly in $(\nu, w) \in \mathbb{R} \times \mathcal{W}$. Therefore, we conclude that $\widehat{F}(\nu, w) \leq F(\nu, w) + o_{\mathbb{P}}(1)$ uniformly in $(\nu, w) \in \mathbb{R} \times \mathcal{W}$.

Arguing similarly, we can also deduce that $\widehat{F}(\nu, w) \geq F(\nu, w) + o_{\mathbb{P}}(1)$ uniformly in $(\nu, w) \in \mathbb{R} \times \mathcal{W}$.

STEP 2: Here we establish that

$$\widehat{t}^* \rightsquigarrow_{\mathbb{P}} \sup_{w \in \mathcal{W}} \Sigma(w)^{-1/2} \big| Z(w) \big| =: t$$

in $\mathbb{R}$. Let $\widetilde{t}^* = \sup_{w \in \mathcal{W}} \Sigma(w)^{-1/2} |\widehat{Z}^*(w)|$. The collection of functions that map elements $z \in \ell^\infty(\mathcal{W})$ to the real line, defined as $\mathcal{G} = \{z \mapsto g_m(z) := m(\|\Sigma^{-1/2}(\cdot)|z(\cdot)|\|_{\mathcal{W}}) : m \in \mathrm{BL}_1(\mathbb{R})\}$, is a subset of $M \cdot \mathrm{BL}_1(\ell^\infty(\mathcal{W}))$, where $M = \sup\{\Sigma^{-1/2}(w) : w \in \mathcal{W}\}$, since $|g_m(z) - g_m(\tilde{z})| \leq M(|z - \tilde{z}|_{\mathcal{W}} \wedge 1)$. Therefore, $d_{\mathrm{BL}}(\widetilde{t}^*, t) \leq M d_{\mathrm{BL}}(\widehat{Z}^*, Z) \to_{\mathbb{P}} 0$, where the latter holds by the definition of $\widehat{Z}^* \rightsquigarrow_{\mathbb{P}} Z$. Moreover, $\widehat{t}^* = \widetilde{t}^* + o_{\mathbb{P}}(1)$, since uniformly over $\mathcal{W}$, $\widehat{\Sigma}^*(w)^{-1/2}|\widehat{Z}^*(w)| = \Sigma(w)^{-1/2}|\widehat{Z}^*(w)| + [\widehat{\Sigma}^*(w)^{-1/2} - \Sigma(w)^{-1/2}]|\widehat{Z}^*(w)| = \Sigma(w)^{-1/2}|\widehat{Z}^*(w)| + o_{\mathbb{P}}(1)$ by part (a) of the lemma. Hence $\widehat{t}^* \rightsquigarrow_{\mathbb{P}} t$ follows from

$$d_{\mathrm{BL}}\big(\widehat{t}^*, t\big) \leq d_{\mathrm{BL}}\big(\widetilde{t}^*, t\big) + E_{M_n}\big[\big|\widetilde{t}^* - \widehat{t}^*\big| \wedge 1\big] \to_{\mathbb{P}} 0,$$

where $E_{M_n}[|\widetilde{t}^* - \widehat{t}^*| \wedge 1] \to_{\mathbb{P}} 0$ by Markov inequality and $E[E_{M_n}[|\widetilde{t}^* - \widehat{t}^*| \wedge 1] \leq E[|o_{\mathbb{P}}(1)| \wedge 1] \to 0$.

STEP 3: Here we show that the limit t is a continuous random variable and $\widehat{t}^*_{1-\alpha} \to_{\mathbb{P}} t_{1-\alpha}$, where $t_{1-\alpha}$ is the $1 - \alpha$ quantile of t. The continuity of the distribution of t follows from Theorem 11.1 of Davydov, Lifshits, and Smorodina (1998) using that $\Sigma(w)$ is nondegenerate over $\mathcal{W}$. The claim $\widehat{t}^*_{1-\alpha} \to_{\mathbb{P}} t_{1-\alpha}$ follows from Step 2, because by the same argument as in Step 1, $\widehat{t}^* \rightsquigarrow_{\mathbb{P}} t$ in $\mathbb{R}$ and t having a continuous distribution function implies convergence of quantiles, namely $\widehat{t}^*_{1-\alpha} \to_{\mathbb{P}} t_{1-\alpha}$. We also remark that this part of the proof is standard; see, for example, Beran (1984).

STEP 4: Here we show claim (b) of the lemma. The argument is standard (e.g., Beran (1984)). Let $\widehat{t} = \sup_{w \in \mathcal{W}} \widehat{\Sigma}^*(w)^{-1/2}|\widehat{Z}(w)|$. By the continuous mapping theorem, $\widehat{t} \rightsquigarrow t$ in $\mathbb{R}$. The event

$$\big\{\Delta(w) \in \big[\widehat{\Delta}^{-*}(w), \widehat{\Delta}^{+*}(w)\big] \text{ for all } w \in \mathcal{W}\big\}$$

is equivalent to the event $\{\widehat{t} \leq \widehat{t}^*_{1-\alpha}\}$. By $\widehat{t}^*_{1-\alpha} \to_{\mathbb{P}} t_{1-\alpha}$ in Step 3, for any $\varepsilon > 0$ and all sufficiently large n, $t_{1-\alpha} - \varepsilon < \widehat{t}^*_{1-\alpha} < t_{1-\alpha} + \varepsilon$. Consequently, $\mathbb{P}\{\widehat{t} \leq t_{1-\alpha} - \varepsilon\} + o(1) \leq \mathbb{P}\{\widehat{t} \leq \widehat{t}^*_{1-\alpha}\} \leq \mathbb{P}\{\widehat{t} \leq t_{1-\alpha} + \varepsilon\} + o(1)$. Taking limits as $n \to \infty$ yields

$$\begin{aligned}\mathbb{P}\{t \leq t_{1-\alpha} - \varepsilon\} &\leq \liminf_{n\to\infty} \mathbb{P}\big\{\widehat{t} \leq \widehat{t}^*_{1-\alpha}\big\} \\ &\leq \limsup_{n\to\infty} \mathbb{P}\big\{\widehat{t} \leq \widehat{t}^*_{1-\alpha}\big\} \leq \mathbb{P}\{t \leq t_{1-\alpha} + \varepsilon\},\end{aligned}$$

where we use that $\widehat{t} \rightsquigarrow t$ in $\mathbb{R}$ and continuity of the distribution function of t. The result follows from taking $\varepsilon \to 0$, using the continuity of the distribution of t, and $\mathbb{P}\{t \leq t_{1-\alpha}\} = 1 - \alpha$. *Q.E.D.*

APPENDIX SB: COMPARISON OF QUANTILE AND DISTRIBUTION REGRESSION: A SIMULATION EXERCISE

In this appendix, we compare quantile and distribution regression as estimators of the conditional and counterfactual distribution functions.

SB.1. *Data Generating Processes*

We calibrate the data generating processes to fit several characteristics of the CPS data sets used in the application in Section 6 of CFM. In particular, we draw the covariate vector X_{88} from the empirical distribution in 1988 (containing 74,661 observations). We consider a model with 8 covariates: a dummy variable for living in metropolitan areas, a dummy variable for part-time work, a dummy variable for not being white, experience, experience squared, a union indicator, education and education squared. These covariates are only a subset of the covariates that we include in the application because we want to consider samples of moderate sizes, which would lead to frequent multicollinearity problems with the 45 covariates used in the application. To estimate a counterfactual distribution, we also draw independent samples of the covariate vector X_{79} from the empirical distribution in 1979 (containing 21,483 observations).

We consider three different data generating processes (DGP) for the conditional distribution of the outcome Y_{88} given X_{88}. We start with a very simple DGP and then show that the conclusions do not change with more realistic models. DGP 1 is the linear location shift model

$$Y_i = X_i'\beta + u_i, \quad i = 1, \dots, n,$$

where the errors u_i are i.i.d. logistically distributed. The coefficient vector β is calibrated to the ordinary least squares (OLS) estimate of the log hourly wage on the covariate vector using the whole 1988 CPS sample. Similarly, the variance of u_i is calibrated to the OLS residual variance.

DGP 2 is the same as DGP 1 except that the errors u_i are drawn i.i.d. from the empirical distribution of the OLS residuals in the whole sample.

DGP 3 is the linear location-scale shift model,

$$Y_i = X_i'\beta + \left(X_i'\delta\right) \cdot u_i, \quad i = 1, \dots, n,$$

where the errors u_i are randomly drawn from the empirical distribution of the OLS residuals in the whole sample rescaled to have variance 1. The coefficient vector β is the same as in the first two DGPs. The coefficient vector δ is calibrated to its OLS estimate obtained by regressing the squared OLS residuals on X_{88}. (We checked that $X_i'\delta > 0$ for $i = 1, \dots, n$.)

Our analysis of the wage distribution in Section 6 reveals a considerable amount of discreteness in the data. A natural mass point is found at the level

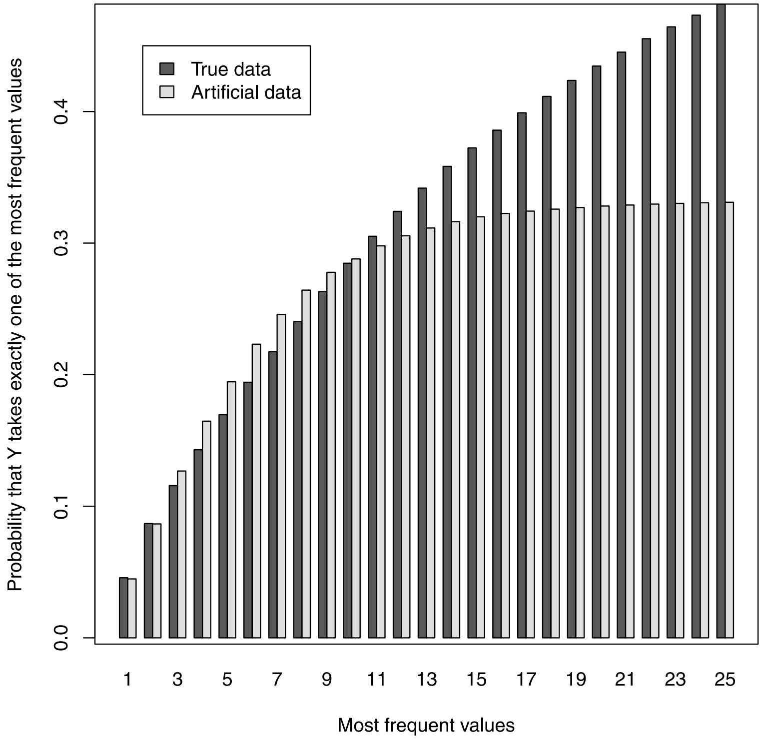

FIGURE SB.1.—Amount of discreteness in the real and simulated data.

of the minimum wage, but other mass points are observed at different levels. Rounded hourly wages are natural focal points in wage negotiation, which could explain the presence of mass points. To analyze the effect of rounding on the quality of the estimation, we consider a version of the three DGPs where wages are rounded to the next dollar with probability $1/3$.

Figure SB.1 compares the amount of rounding present in the wage data in 1988 and the amount of rounding implied by our data generating process.[1] It appears that the simulated data match very well the probability that the wage takes exactly one of the ten most frequent values. For instance, the observed wages take one of the two most frequent values with probability 8.7%, while this probability is 8.6% for the simulated data. Our data generating process does not round the dependent variable enough when we consider the values that are not among the ten most frequent values. In this sense, the discreteness implied by our DGPs is more modest than that observed in the data.

[1]We plot these probabilities for DGP 1. The probabilities are very similar for the two other DGPs.

SB.2. *Estimands of Interest and Estimators*

The first estimand of interest is the conditional distribution function of Y_{88} given X_{88},

$$F_{Y_{88}|X_{88}}(y|x) \quad \text{for} \quad (y,x) \in \mathcal{YX},$$

where $\mathcal{YX}$ is the joint support of Y_{88} and X_{88}. We are interested in the conditional distribution because this quantity is often of direct economic interest and it is the first step in the estimation of the counterfactual distributions. Our second and main estimand is the counterfactual distribution obtain by integrating this distribution over the empirical marginal distribution of X in 1979:

$$F_{Y\langle 88|79\rangle}(y) = \int_{\mathcal{X}} F_{Y_{88}|X_{88}}(y|x)\, dF_{X_{79}}(x).$$

We implement the distribution regression estimator using the logistic link function. For 500 different cutoff values y located at equidistant marginal quantiles of Y_{88}, we regress $1(Y_{88} \leq y)$ on X_{88} using a linear logit regression. The estimated conditional distribution is obtained as

$$\widehat{F}^{\mathrm{DR}}_{Y_{88}|X_{88}}(y|x) = \Lambda\big(x'\widehat{\beta}(y)\big),$$

where $\Lambda(\cdot)$ is the logistic distribution function. Since this estimated conditional distribution function may be nonmonotonic in y, we also apply the monotonization method of Chernozhukov, Fernández-Val, and Galichon (2010) based on rearrangement. The rearranged c.d.f. is[2]

$$\widehat{F}^{*}_{Y_{88}|X_{88}}(y|x) = \inf\left\{u : \int_0^{\infty} 1\big(\widehat{F}^{\mathrm{DR}}_{Y_{88}|X_{88}}(t|x) \leq u\big)\, dt \geq y\right\}.$$

To implement the quantile estimator, we estimate 500 linear quantile regressions of Y_{88} on X_{88} (Koenker and Bassett (1978)) and obtain

$$\widehat{Q}_{Y_{88}|X_{88}}(\tau|x) = x'\widehat{\beta}(\tau) \quad \text{for} \quad \tau = 0.001, \dots, 0.999.$$

We invert this estimated quantile function to obtain the c.d.f. using

$$\widehat{F}^{\mathrm{QR}}_{Y_{88}|X_{88}}(y|x) = \frac{1}{1000} + \int_{1/1000}^{999/1000} 1\big(\widehat{Q}_{Y_{88}|X_{88}}(u|x) \leq y\big)\, du.$$

The counterfactual distribution of interest, $F_{Y\langle 88|79\rangle}(y)$, is estimated by averaging the estimator of the conditional distribution in 1988 over the covariate

[2]This expression is simplified by the fact that Y is nonnegative.

distribution in 1979, that is,

$$\widehat{F}_{Y\langle 88|79\rangle}(y) = \frac{1}{n}\sum_{i=1}^{n}\widehat{F}_{Y_{88}|X_{88}}(y|X_{79,i}),$$

where $\widehat{F}_{Y_{88}|X_{88}}$ is $\widehat{F}^{\mathrm{DR}}_{Y_{88}|X_{88}}$, $\widehat{F}^{*}_{Y_{88}|X_{88}}$, or $\widehat{F}^{\mathrm{QR}}_{Y_{88}|X_{88}}$. Thus, two independent samples of size n are used to estimate this counterfactual distribution: n observations drawn from (X_{88}, Y_{88}) to obtain $\widehat{F}_{Y_{88}|X_{88}}$ and n observations drawn from X_{79} to obtain $(X_{79,1}, \dots, X_{79,n})$

SB.3. *Results*

We measure the performance of the estimators by the integrated mean squared error (MSE) and the integrated Anderson–Darling weighted MSE. The MSEs for the counterfactual and conditional distributions are

$$\int_{\mathcal{Y}}\big(\widehat{F}_{Y\langle 88|79\rangle}(y) - F_{Y\langle 88|79\rangle}(y)\big)^2\,dF_{Y\langle 88|79\rangle}(y),$$

$$\int_{\mathcal{X}}\int_{\mathcal{Y}}\big(\widehat{F}_{Y_{88}|X_{88}}(y|x) - F_{Y_{88}|X_{88}}(y|x)\big)^2\,dF_{Y_{88}|X_{88}}(y|x)\,dF_{X_{88}}(x).$$

Similarly, the Anderson–Darling weighted MSEs are

$$\int_{\mathcal{Y}}\frac{(\widehat{F}_{Y\langle 88|79\rangle}(y) - F_{Y\langle 88|79\rangle}(y))^2}{F_{Y\langle 88|79\rangle}(y)(1 - F_{Y\langle 88|79\rangle}(y))}\,dF_{Y\langle 88|79\rangle}(y),$$

$$\int_{\mathcal{X}}\int_{\mathcal{Y}}\frac{(\widehat{F}_{Y_{88}|X_{88}}(y|x) - F_{Y_{88}|X_{88}}(y|x))^2}{F_{Y_{88}|X_{88}}(y|x)(1 - F_{Y_{88}|X_{88}}(y|x))}\,dF_{Y_{88}|X_{88}}(y|x)\,dF_{X_{88}}(x).$$

These integrals are approximated numerically by averages over 1000 points of x drawn randomly from the distribution of X_{88} and over 100 points of y corresponding to 100 quantiles of Y_{88} on a uniform grid between 0 and 1.

Tables SB.I–SB.VI report the results. Since the relative efficiency of the estimators does not vary when we consider the MSE or the weighted MSE, we will comment in detail on the MSE results only. This similarity means that the relative performance of the estimators does not change much when we move from the center of the distribution to the tails.

Tables SB.I and SB.II provide the detailed results for DGP 1 in the absence and in the presence of rounding for 100, 1000, and 10,000 observations. When the dependent variable is continuous, the quantile regression estimator produces the most accurate estimates for both the counterfactual and the conditional distributions, and for all sample sizes considered. In this DGP, all estimators are correctly specified. The distribution regression estimator performs

TABLE SB.I

SQUARE ROOT MSE OF THE DISTRIBUTION AND QUANTILE ESTIMATORS FOR THE DGP 1[a]

Estimand	No. of Obs.	No Rounding $\hat{F}_Y^{DR}$	No Rounding $\hat{F}_Y^{*}$	No Rounding $\hat{F}_Y^{QR}$	Rounding With Pr = 1/3 $\hat{F}_Y^{DR}$	Rounding With Pr = 1/3 $\hat{F}_Y^{*}$	Rounding With Pr = 1/3 $\hat{F}_Y^{QR}$
Counterfactual distribution	100	5.09	4.73	3.96	5.53	5.16	4.50
	1000	1.42	1.41	1.19	1.54	1.54	1.86
	10,000	0.46	0.46	0.39	0.50	0.50	1.40
Conditional distribution	100	18.76	16.11	12.41	20.35	17.56	13.68
	1000	5.11	4.94	3.86	5.55	5.38	4.66
	10,000	1.57	1.57	1.23	1.72	1.71	2.44

[a]All numbers are multiplied by 1000.

worse because it binarizes the information contained in the value of the dependent variable. Rearranging the estimates always improves the performance of the distribution regression estimator, although this improvement vanishes when the sample size increases.

The right panels of Tables SB.I and SB.II provide the results in the presence of discrete mass points in the wage distribution. Rounding one-third of the wages to the next dollar is enough to reverse the order between the MSE of the quantile and distribution regression estimators, when the number of observations is above 100 for the counterfactual distribution and when the number of observations is above 1000 for the conditional distribution. The rounding introduces a misspecification of the linear quantile regression model while it does not affect the validity of the logit model for the conditional distribution. Thus, the linear quantile regression estimator is a better estimator when the dependent variable is continuous, but a realistic amount of rounding is enough to reverse the results.

TABLE SB.II

SQUARE ROOT ANDERSON–DARLING WEIGHTED MSE OF THE DISTRIBUTION AND QUANTILE ESTIMATORS FOR THE DGP 1[a]

Estimand	No. of Obs.	No Rounding $\hat{F}_Y^{DR}$	No Rounding $\hat{F}_Y^{*}$	No Rounding $\hat{F}_Y^{QR}$	Rounding With Pr = 1/3 $\hat{F}_Y^{DR}$	Rounding With Pr = 1/3 $\hat{F}_Y^{*}$	Rounding With Pr = 1/3 $\hat{F}_Y^{QR}$
Counterfactual distribution	100	14.80	11.55	9.64	16.35	12.60	10.95
	1000	3.48	3.44	2.86	3.78	3.75	4.36
	10,000	1.12	1.12	0.93	1.22	1.22	3.22
Conditional distribution	100	54.51	38.40	30.39	59.87	42.03	33.58
	1000	12.85	11.85	9.46	14.05	12.95	11.28
	10,000	3.78	3.76	3.03	4.14	4.12	5.64

[a]All numbers are multiplied by 1000.

TABLE SB.III

SQUARE ROOT MSE OF THE DISTRIBUTION AND QUANTILE ESTIMATORS FOR THE DGP 2[a]

		No Rounding			Rounding With Pr = 1/3		
Estimand	No. of Obs.	$\hat{F}_Y^{DR}$	$\hat{F}_Y^{*}$	$\hat{F}_Y^{QR}$	$\hat{F}_Y^{DR}$	$\hat{F}_Y^{*}$	$\hat{F}_Y^{QR}$
Counterfactual distribution	100	5.04	4.69	3.99	5.46	5.09	4.5
	1000	1.44	1.43	1.19	1.56	1.55	1.8
	10,000	0.46	0.46	0.39	0.5	0.5	1.33
Conditional distribution	100	18.62	15.9	12.29	20.24	17.38	13.54
	1000	5.11	4.93	3.87	5.56	5.38	4.62
	10,000	1.62	1.61	1.23	1.77	1.76	2.32

[a]All numbers are multiplied by 1000.

One may argue that DGP 1 favors the distribution regression estimator because the error terms are logistically distributed. The results for DGP 2 reported in Tables SB.III and SB.IV show that this ingredient is not crucial. Replacing the logistic distribution with the empirical distribution of the OLS residuals in the whole population does not change qualitatively the results.

DGP 3 relaxes the independence assumption between regressors and errors by introducing linear multiplicative heteroscedasticity. In the absence of rounding, the linear quantile regression estimator is still correctly specified, while the logit regression is doubly misspecified (nonlogistic error terms and heteroscedasticity). In this sense, DGP 3 favors the quantile regression estimator. Despite this, the results for the counterfactual distribution in Tables SB.V and SB.VI show again that the distribution regression estimator performs better than quantile regression in the presence of rounding and at least 1000 observations. The difference is that now quantile regression is a better estimator of the conditional distribution even in the presence of rounding. In

TABLE SB.IV

SQUARE ROOT ANDERSON–DARLING WEIGHTED MSE OF THE DISTRIBUTION AND QUANTILE ESTIMATORS FOR THE DGP 2[a]

		No Rounding			Rounding With Pr = 1/3		
Estimand	No. of Obs.	$\hat{F}_Y^{DR}$	$\hat{F}_Y^{*}$	$\hat{F}_Y^{QR}$	$\hat{F}_Y^{DR}$	$\hat{F}_Y^{*}$	$\hat{F}_Y^{QR}$
Counterfactual distribution	100	14.51	11.45	9.66	16.02	12.48	10.95
	1000	3.51	3.47	2.87	3.83	3.79	4.28
	10,000	1.12	1.12	0.94	1.23	1.23	3.11
Conditional distribution	100	54.05	38.00	30.20	59.67	41.73	33.37
	1000	12.84	11.87	9.49	14.11	13.01	11.23
	10,000	3.88	3.87	3.04	4.28	4.26	5.43

[a]All numbers are multiplied by 1000.

TABLE SB.V

SQUARE ROOT MSE OF THE DISTRIBUTION AND QUANTILE ESTIMATORS FOR THE DGP 3[a]

		No Rounding			Rounding With Pr = 1/3		
Estimand	No. of Obs.	$\hat{F}_Y^{DR}$	$\hat{F}_Y^{*}$	$\hat{F}_Y^{QR}$	$\hat{F}_Y^{DR}$	$\hat{F}_Y^{*}$	$\hat{F}_Y^{QR}$
Counterfactual distribution	100	4.98	4.63	3.98	5.43	5.06	4.52
	1000	1.43	1.42	1.18	1.55	1.54	1.86
	10,000	0.5	0.5	0.39	0.55	0.55	1.42
Conditional distribution	100	18.94	16.16	12.27	20.62	17.68	13.56
	1000	5.75	5.58	3.9	6.19	6.03	4.74
	10,000	3.02	3.02	1.29	3.17	3.17	2.54

[a] All numbers are multiplied by 1000.

our application, the main objects of interest are counterfactual distributions, we have more than 20,000 observations in one period and more than 70,000 in the other, and the amount of rounding is at least as high as in the simulations. Therefore, even if the true DGP corresponded to DGP 3, we would prefer the distribution regression approach.

Since the presence of mass points in the distribution of the dependent variable penalizes the relative performance of the quantile regression estimator, we considered applying a small random noise to the dependent variable (also called dithering or jittering). This noise artificially restores the continuity of the distribution. The results, however, showed no improvement (even a small deterioration) in the MSE of the quantile regression estimator. On the other hand, it may be a way to restore the validity of the inference procedures.

TABLE SB.VI

SQUARE ROOT ANDERSON–DARLING WEIGHTED MSE OF THE DISTRIBUTION AND QUANTILE ESTIMATORS FOR THE DGP 3[a]

		No Rounding			Rounding With Pr = 1/3		
Estimand	No. of Obs.	$\hat{F}_Y^{DR}$	$\hat{F}_Y^{*}$	$\hat{F}_Y^{QR}$	$\hat{F}_Y^{DR}$	$\hat{F}_Y^{*}$	$\hat{F}_Y^{QR}$
Counterfactual distribution	100	14.35	11.27	9.63	15.77	12.33	10.96
	1000	3.46	3.42	2.85	3.76	3.73	4.40
	10,000	1.19	1.19	0.93	1.29	1.29	3.30
Conditional distribution	100	54.95	39.01	30.54	60.41	42.80	33.76
	1000	14.64	13.74	9.67	15.76	14.81	11.58
	10,000	7.66	7.65	3.31	7.98	7.97	6.00

[a] All numbers are multiplied by 1000.

SB.4. *Conclusion*

We draw two main lessons from this section. First, the quantile regression estimator is more accurate than the distribution regression estimator for data generating processes that satisfy the assumptions of both models. Our results agree with the recent simulations of Koenker (2010). Second, introducing a realistic amount of discreteness of the dependent variable is enough to revert the results. The distribution regression approach is naturally robust to such a ubiquitous phenomenon, while quantile regression is not.

Quantile and distribution regression make different parametric assumptions. It is, therefore, easy to find data generating processes for which one estimator dominates the other because of misspecification. This was not the goal of this simulation exercise, but this is an empirical question, the response to which changes from one case to the other. In our application, for instance, we think that it is important to allow for different coefficients below and above the minimum wage. Moreover, the misspecification tests reported in Rothe and Wied (2012) reject the quantile regression, but not the distribution regression estimator for a similar data set.

On the bright side, both distribution regression (independently of the link function) and quantile regression give numerically identical results in saturated models. Therefore, the choice between distribution or quantile regression becomes immaterial if we have a flexible enough specification. In our application, while the results are significantly different statistically, their economic interpretation remains extremely similar.

APPENDIX SC: ADDITIONAL EMPIRICAL RESULTS

Table SC.I contains descriptive statistics for the data sets used in CFM and in this appendix. Between 1979 and 1988, the average real wage decreased for men while it increased for women. The level of potential experience decreased because of the entry of the baby-boom generation into the labor market and because of longer education. Educational attainment increased clearly over the period. As is well known, de-unionization was important with a 11 percent fall in union members for men and 5 percent for women. The rise of the service sector is another remarkable change that took place during this period.

Figure SC.1 presents robustness checks with respect to the link functions used for the distribution regression estimation. The differences between the estimates obtained with the logistic, normal, uniform (linear probability model), Cauchy, and complementary log–log link functions are so modest that the lines are almost indistinguishable.

Table SC.II and Figure SC.2 present robustness checks with respect to the wage mechanism below the minimum wage. The assumptions about the minimum wage are particularly delicate, since the mechanism that generates wages strictly below this level is not clear; it could be measurement error, noncoverage, or noncompliance with the law. To check the robustness of the results

TABLE SC.I

SUMMARY STATISTICS[a]

Variable	Males 1979	Males 1988	Females 1979	Females 1988
No. of obs.	21,483	74,661	16,911	70,089
Wage	7.24 (3.71)	7.01 (4.32)	4.75 (2.39)	5.05 (2.99)
Education	12.44 (2.98)	12.95 (2.80)	12.47 (2.48)	13.02 (2.44)
Experience	18.69 (13.55)	18.02 (12.25)	18.28 (13.66)	17.81 (12.39)
Union	31.77 (46.56)	20.58 (40.43)	17.21 (37.75)	12.65 (33.24)
Primary sector	4.59 (20.93)	3.81 (19.14)	1.20 (10.89)	1.06 (10.22)
Secondary sector	39.99 (48.99)	34.90 (47.67)	19.06 (39.28)	15.20 (35.90)
Tertiary sector	55.42 (49.71)	61.29 (48.71)	79.74 (40.19)	83.74 (36.90)
Part-time	6.93 (25.39)	10.55 (30.73)	22.20 (41.56)	25.30 (43.47)
Non-white	9.79 (29.72)	11.30 (31.66)	13.09 (33.73)	13.68 (34.37)
SMSA	59.17 (49.15)	73.09 (44.35)	60.84 (48.81)	73.72 (44.02)

[a]Mean of selected variables. The standard deviations are reported in parentheses. When the variable is binary, the results are reported in percentage.

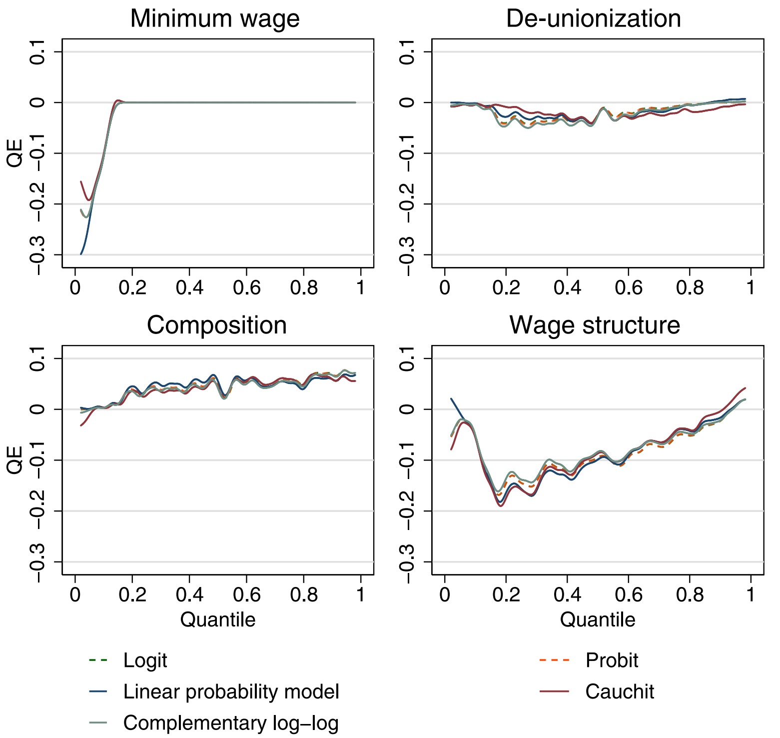

FIGURE SC.1.—Comparison of the distribution regression estimates of the quantile effects based on five different link functions: logistic, normal, uniform, Cauchy and complementary log-log.

TABLE SC.II

DECOMPOSING CHANGES IN MEASURES OF WAGE DISPERSION: MEN, CENSORED MODELS[a]

		Effect of			
Statistic	Total Change	Minimum Wage	Unions	Individual Attributes	Coefficients
CDR:					
Standard deviation	8.2 (0.3)	3.3 (0.0)	0.6 (0.0)	1.9 (0.2)	2.4 (0.2)
		40.7 (1.4)	7.9 (0.5)	22.5 (1.8)	28.9 (2.4)
90–10	21.5 (1.0)	11.2 (0.1)	0.0 (0.0)	9.2 (0.8)	1.1 (1.3)
		52.1 (2.4)	0.0 (0.1)	42.6 (4.4)	5.3 (5.9)
50–10	11.3 (1.4)	11.2 (0.1)	−2.0 (1.0)	5.1 (0.4)	−3.1 (1.1)
		99.6 (14.1)	−17.9 (11.2)	45.5 (8.3)	−27.2 (14.0)
90–50	10.2 (1.2)	0.0 (0.0)	2.0 (1.0)	4.0 (0.8)	4.2 (1.1)
		0.0 (0.0)	19.7 (8.4)	39.3 (8.8)	41.0 (9.8)
75–25	15.4 (1.1)	0.0 (0.0)	4.1 (1.0)	0.3 (1.3)	11.1 (1.2)
		0.0 (0.0)	26.5 (6.2)	1.7 (8.6)	71.8 (8.7)
95–5	36.4 (2.1)	26.4 (0.7)	0.0 (0.6)	8.5 (1.1)	1.4 (1.5)
		72.7 (3.8)	0.0 (1.5)	23.4 (2.7)	3.9 (4.0)
Gini coefficient	4.2 (0.1)	1.6 (0.0)	0.4 (0.0)	0.3 (0.1)	1.8 (0.1)
		37.9 (1.1)	10.7 (0.5)	7.1 (1.6)	44.2 (1.6)
CQR:					
Standard deviation	9.0 (0.3)	4.1 (0.0)	0.3 (0.0)	1.8 (0.1)	2.8 (0.2)
		45.3 (1.5)	3.2 (0.5)	20.0 (1.6)	31.4 (2.2)
90–10	22.3 (1.1)	14.2 (0.4)	−0.5 (0.1)	7.2 (0.4)	1.4 (1.1)
		63.6 (3.4)	−2.2 (0.6)	32.3 (2.8)	6.2 (5.1)
50–10	9.5 (0.9)	14.2 (0.4)	−1.8 (0.1)	4.6 (0.4)	−7.5 (0.9)
		149.2 (16.7)	−18.7 (3.0)	48.0 (9.0)	−78.5 (21.6)
90–50	12.7 (0.7)	0.0 (0.0)	1.3 (0.1)	2.6 (0.3)	8.8 (0.5)
		0.0 (0.0)	10.1 (1.0)	20.6 (2.4)	69.3 (2.5)
75–25	12.7 (0.6)	0.0 (0.0)	1.7 (0.1)	2.0 (0.4)	9.1 (0.5)
		0.0 (0.0)	13.0 (1.2)	15.5 (3.0)	71.4 (3.1)
95–5	39.2 (0.8)	30.6 (0.0)	−0.5 (0.1)	7.4 (0.5)	1.6 (0.8)
		78.1 (1.8)	−1.2 (0.3)	18.9 (1.2)	4.2 (2.1)
Gini coefficient	4.5 (0.1)	1.9 (0.0)	0.3 (0.0)	0.3 (0.1)	2.1 (0.1)
		42.2 (1.1)	5.9 (0.4)	6.1 (1.4)	45.8 (1.4)

[a]CDR and CQR denote censored distribution regression and censored quantile regression, respectively. The censored logit distribution regression and the censored linear quantile regression estimators have been applied. All numbers are in percentages. Bootstrapped standard errors with 100 repetitions are given in parentheses. The second line in each cell indicates the percentage of total variation.

to the DFL assumptions about the minimum wage and to our semiparametric model of the conditional distribution, we reestimate the decomposition using censored linear quantile regression and censored distribution regression with a logit link, censoring the wage data at the level of the minimum wage.

FIGURE SC.2.—Comparison of the distribution regression, censored distribution regression and censored quantile regression estimates of the quantile effects.

For censored quantile regression, we use Powell's (1986) censored quantile regression estimated using Chernozhukov and Hong's (2002) algorithm. For censored distribution regression, we simply censor to zero the distribution regression estimates of the conditional distributions below the minimum wage and recompute the functionals of interest. Overall, we find that the results are very similar for the quantile and distribution regressions, and they are not very sensitive to the censoring. Table SC.III shows that reversing the order of the factors to (i) labor force composition, (ii) de-unionization, (iii) minimum wage, and (iv) wage structure has little qualitative effect.

We present our results for female workers in Tables SC.IV and SC.V as well as Figures SC.3–SC.7. Table SC.IV reports the decomposition of the changes in various measures of wage dispersion between 1979 and 1988 estimated using logit distribution regressions. Table SC.V reports the results of the same decomposition estimated using the censored distribution regression and the censored quantile regression estimators. Figures SC.3–SC.5 refine these results by presenting estimates and 95 percent simultaneous confidence bands for quantile, distribution, and Lorenz effects. The procedures used are exactly the same as were used to produce the tables and figures for men. Figures SC.6

TABLE SC.III

REVERSING THE ORDER OF THE DECOMPOSITION[a]

Statistic	Total Change	Effect of: Individual Attributes	Unions	Minimum Wage	Coefficients
Men:					
Standard deviation	8.0 (0.3)	0.9 (0.2) 11.4 (2.3)	1.5 (0.1) 19.2 (1.0)	2.8 (0.1) 35.3 (1.5)	2.7 (0.3) 34.1 (2.5)
90–10	21.5 (1.1)	0.4 (1.2) 1.8 (5.7)	8.8 (1.2) 40.7 (5.7)	11.2 (0.3) 52.1 (3.1)	1.1 (1.3) 5.3 (5.9)
50–10	11.3 (1.4)	2.5 (1.5) 21.8 (12.6)	0.7 (1.2) 5.8 (11.6)	11.2 (0.3) 99.6 (15.1)	−3.1 (1.1) −27.1 (15.2)
90–50	10.2 (1.2)	−2.1 (1.1) −20.1 (11.9)	8.1 (0.8) 79.1 (10.6)	0.0 (0.0) 0.0 (0.0)	4.2 (1.3) 41.0 (10.3)
75–25	15.4 (1.0)	6.4 (1.2) 41.6 (6.6)	−2.1 (1.3) −13.4 (9.0)	0.0 (0.0) 0.0 (0.0)	11.1 (1.1) 71.8 (8.3)
95–5	33.0 (2.0)	2.6 (1.5) 7.9 (4.1)	5.9 (1.1) 17.9 (3.5)	23.0 (0.8) 69.9 (4.0)	1.4 (1.3) 4.3 (4.0)
Gini coefficient	4.1 (0.1)	−0.3 (0.1) −6.8 (2.4)	1.0 (0.0) 24.1 (1.0)	1.3 (0.0) 32.4 (1.4)	2.0 (0.1) 50.4 (2.0)
Women:					
Standard deviation	10.9 (0.4)	4.5 (0.2) 41.1 (1.9)	0.0 (0.0) −0.2 (0.2)	4.0 (0.2) 36.5 (1.7)	2.5 (0.3) 22.6 (2.6)
90–10	39.8 (1.4)	11.2 (0.7) 28.2 (1.6)	0.0 (0.4) 0.0 (1.0)	27.2 (0.3) 68.4 (2.4)	1.3 (1.2) 3.4 (2.8)
50–10	33.0 (0.7)	7.9 (0.9) 24.1 (2.5)	−0.8 (0.5) −2.4 (1.7)	27.2 (0.3) 82.6 (2.2)	−1.4 (0.8) −4.3 (2.4)
90–50	6.8 (1.4)	3.3 (0.8) 47.9 (13.3)	0.8 (0.5) 11.7 (7.0)	0.0 (0.0) 0.0 (0.0)	2.8 (1.4) 40.3 (10.1)
75–25	12.8 (0.9)	2.8 (0.8) 22.0 (5.4)	0.0 (0.5) 0.0 (3.8)	5.5 (0.1) 43.0 (3.1)	4.5 (0.8) 35.0 (5.1)
95–5	38.8 (1.7)	17.4 (0.9) 44.8 (2.7)	0.0 (0.3) 0.0 (0.9)	9.0 (2.7) 23.3 (6.4)	12.4 (2.4) 31.9 (6.1)
Gini coefficient	5.1 (0.2)	0.7 (0.1) 13.9 (1.9)	0.1 (0.0) 1.1 (0.3)	2.8 (0.1) 55.9 (2.4)	1.5 (0.2) 29.1 (2.5)

[a]The logit distribution regression model has been applied. All numbers are in percentages. Bootstrapped standard errors with 100 repetitions are given in parentheses. The second line in each cell indicates the percentage of total variation.

and SC.7 show that these results are not sensitive to the choice of the estimator of the conditional distribution.

Most of the patterns for women are similar to men, although there are a few interesting differences in the strength of the effects that explain the changes in inequality. First, de-unionization is virtually irrelevant for women. This is

TABLE SC.IV

DECOMPOSING CHANGES IN MEASURES OF WAGE DISPERSION: WOMEN, DISTRIBUTION REGRESSION[a]

Statistic	Total Change	Effect of: Minimum Wage	Unions	Individual Attributes	Coefficients
Standard deviation	10.9 (0.4)	3.8 (0.1) 34.9 (1.6)	0.4 (0.0) 3.2 (0.4)	4.7 (0.2) 42.8 (1.6)	2.1 (0.3) 19.1 (2.5)
90–10	39.8 (1.1)	23.0 (0.3) 57.9 (1.5)	0.9 (0.5) 2.3 (1.2)	14.5 (0.7) 36.4 (1.6)	1.3 (0.9) 3.4 (2.1)
50–10	33.0 (0.8)	23.0 (0.3) 69.9 (1.9)	0.0 (0.1) 0.0 (0.4)	11.3 (0.4) 34.4 (1.4)	−1.4 (0.8) −4.3 (2.8)
90–50	6.8 (1.3)	0.0 (0.0) 0.0 (0.0)	0.9 (0.5) 13.6 (7.0)	3.1 (0.7) 46.0 (11.6)	2.8 (1.3) 40.3 (9.7)
75–25	12.8 (0.9)	0.0 (0.0) 0.0 (0.0)	0.0 (0.5) 0.0 (3.5)	8.3 (0.2) 65.0 (5.1)	4.5 (0.8) 35.0 (4.3)
95–5	38.8 (1.6)	16.8 (0.6) 43.2 (2.4)	0.7 (0.7) 1.9 (2.0)	16.4 (2.0) 42.1 (4.9)	5.0 (2.0) 12.8 (5.1)
Gini coefficient	5.1 (0.2)	2.4 (0.1) 47.3 (2.0)	0.2 (0.0) 3.5 (0.4)	1.3 (0.1) 24.9 (1.2)	1.2 (0.1) 24.2 (2.3)

[a]The logit distribution regression model has been applied. All numbers are in percentages. Bootstrapped standard errors with 100 repetitions are given in parentheses. The second line in each cell indicates the percentage of total variation.

due to the fact that the proportion of unionized female workers has always been small. The decline in the unionization rate is smaller for females than for males in absolute value. In addition, unions do not compress the conditional female wage distribution, while they do reduce the conditional variance of the male wage distribution. Second, the decrease in the real value of the minimum wage explains a larger increase in wage inequality for women than for men. The reason is that the proportion of workers at or below the minimum wage is higher for female than for male workers. Thus, mechanically, women will be more affected by a decrease in the value of the minimum wage. Third, changes in individual attributes generally have a more important effect than changes in the wage structure, which is not true for men. The robustness checks using the censored distribution and quantile models as well as the different link functions for the distribution regression estimators confirm these results.

APPENDIX SD: VARIANCE DECOMPOSITION INTO BETWEEN-GROUP AND WITHIN-GROUP COMPONENTS

By the law of total variance, we can decompose the variance as

$$\mathrm{Var}(Y) = \mathrm{Var}\big[E(Y|X)\big] + E\big[\mathrm{Var}(Y|X)\big],$$

TABLE SC.V

DECOMPOSING CHANGES IN MEASURES OF WAGE DISPERSION: WOMEN, CENSORED MODELS[a]

Statistic	Total Change	Effect of Minimum Wage	Effect of Unions	Effect of Individual Attributes	Effect of Coefficients
CDR:					
Standard deviation	12.7 (0.4)	5.6 (0.0)	0.3 (0.0)	5.1 (0.2)	1.7 (0.3)
		44.1 (1.2)	2.2 (0.3)	39.9 (1.4)	13.8 (2.2)
90–10	43.2 (1.1)	26.4 (0.3)	0.9 (0.5)	14.5 (0.7)	1.3 (0.9)
		61.2 (1.5)	2.2 (1.1)	33.5 (1.5)	3.1 (1.9)
50–10	36.4 (0.8)	26.4 (0.3)	0.0 (0.1)	11.3 (0.4)	−1.4 (0.8)
		72.7 (1.8)	0.0 (0.4)	31.2 (1.2)	−3.9 (2.5)
90–50	6.8 (1.3)	0.0 (0.0)	0.9 (0.5)	3.1 (0.7)	2.8 (1.3)
		0.0 (0.0)	13.6 (7.0)	46.0 (11.6)	40.3 (9.7)
75–25	12.8 (0.9)	0.0 (0.0)	0.0 (0.5)	8.3 (0.2)	4.5 (0.8)
		0.0 (0.0)	0.0 (3.5)	65.0 (5.1)	35.0 (4.3)
95–5	52.7 (1.2)	30.6 (0.0)	0.7 (0.3)	16.7 (0.8)	4.7 (1.1)
		58.1 (1.5)	1.4 (0.6)	31.6 (1.4)	8.8 (0.2)
Gini coefficient	6.4 (0.1)	3.6 (0.0)	0.1 (0.0)	1.7 (0.1)	0.9 (0.1)
		57.1 (1.3)	2.0 (0.2)	27.1 (1.0)	13.8 (1.8)
CQR:					
Standard deviation	12.9 (0.3)	6.2 (0.0)	0.3 (0.1)	4.5 (0.2)	1.8 (0.3)
		48.2 (1.3)	2.6 (0.4)	35.2 (1.5)	13.9 (2.2)
90–10	48.5 (0.9)	30.6 (0.0)	0.7 (0.2)	14.6 (0.6)	2.5 (0.9)
		63.2 (1.2)	1.5 (0.3)	30.2 (1.1)	5.1 (1.7)
50–10	37.2 (0.6)	30.6 (0.0)	−0.3 (0.1)	10.9 (0.5)	−4.1 (0.5)
		82.3 (1.3)	−0.7 (0.2)	29.4 (1.3)	−10.9 (1.6)
90–50	11.3 (0.8)	0.0 (0.0)	1.0 (0.1)	3.7 (0.4)	6.5 (0.8)
		0.0 (0.0)	9.1 (1.1)	32.8 (3.8)	58.1 (4.0)
75–25	15.2 (0.8)	0.0 (0.0)	0.8 (0.1)	11.9 (0.6)	2.5 (0.8)
		0.0 (0.0)	5.6 (0.7)	78.1 (4.9)	16.4 (4.9)
95–5	50.1 (1.2)	30.6 (0.0)	1.0 (0.2)	15.1 (0.7)	3.4 (1.1)
		61.1 (1.4)	2.0 (0.4)	30.2 (1.2)	6.7 (2.0)
Gini coefficient	6.5 (0.1)	4.0 (0.0)	0.1 (0.0)	1.5 (0.1)	0.9 (0.1)
		60.9 (1.3)	2.1 (0.3)	23.1 (1.2)	13.9 (1.8)

[a] CDR and CQR are as defined in Table SC.II. All numbers are in percentages. Bootstrapped standard errors with 100 repetitions are given in parentheses. The second line in each cell indicates the percentage of total variation. The censored distribution regression estimator with logistic link and the censored linear quantile regression estimator have been applied.

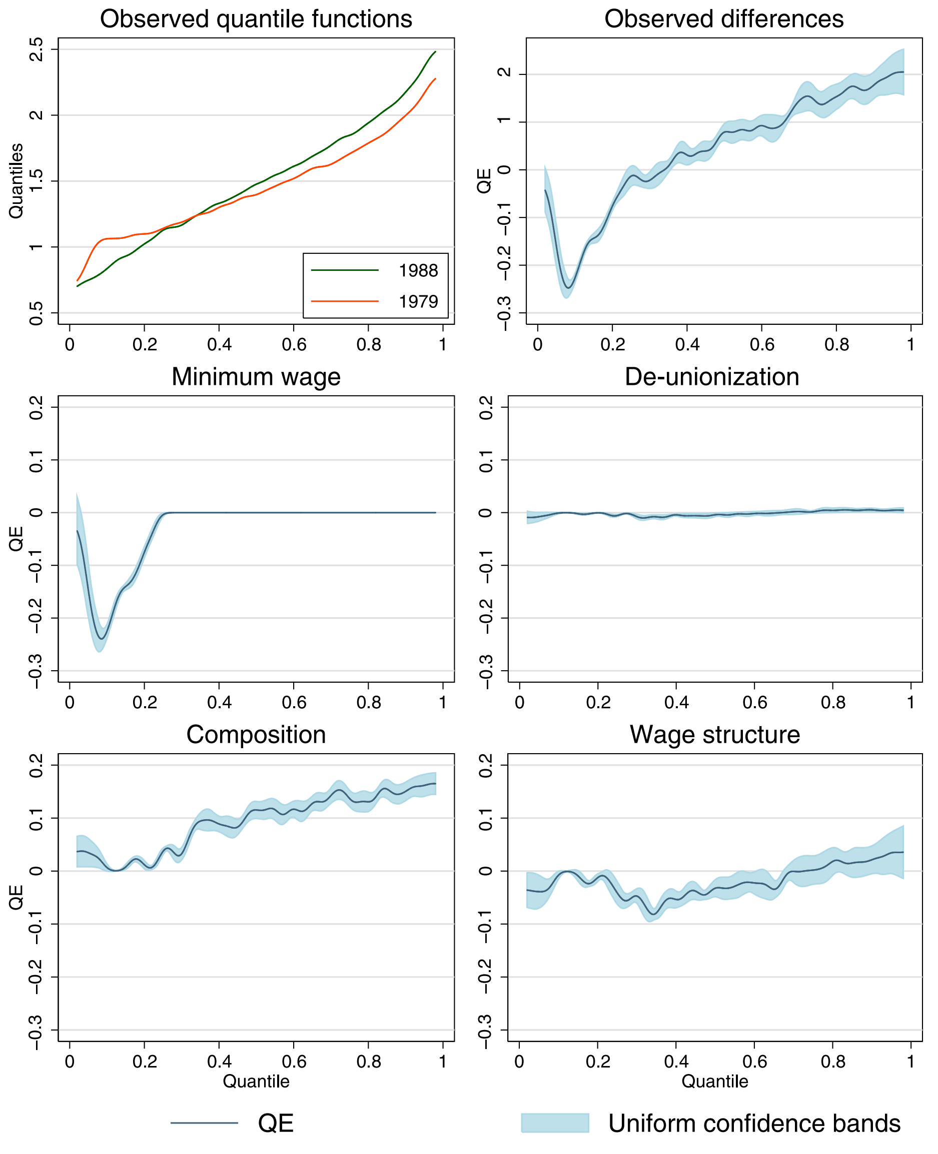

FIGURE SC.3.—Observed quantile functions, observed differences between the quantile functions and their decomposition into four quantile effects. The 95% simultaneous confidence bands were obtained by empirical bootstrap with 100 repetitions. Results for women.

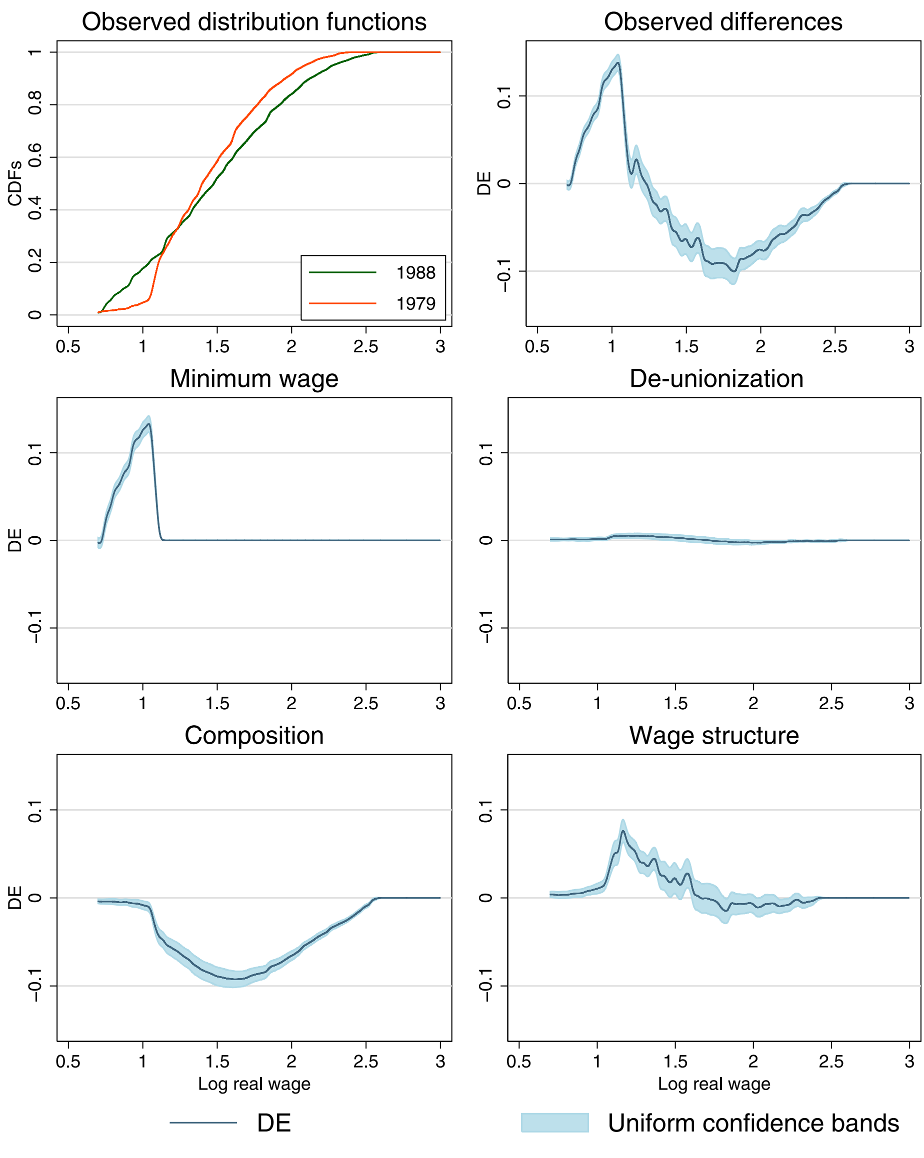

FIGURE SC.4.—Observed distribution functions, observed differences between the distribution functions and their decomposition into four distribution effects. The 95% simultaneous confidence bands were obtained by empirical bootstrap with 100 repetitions. Results for women.

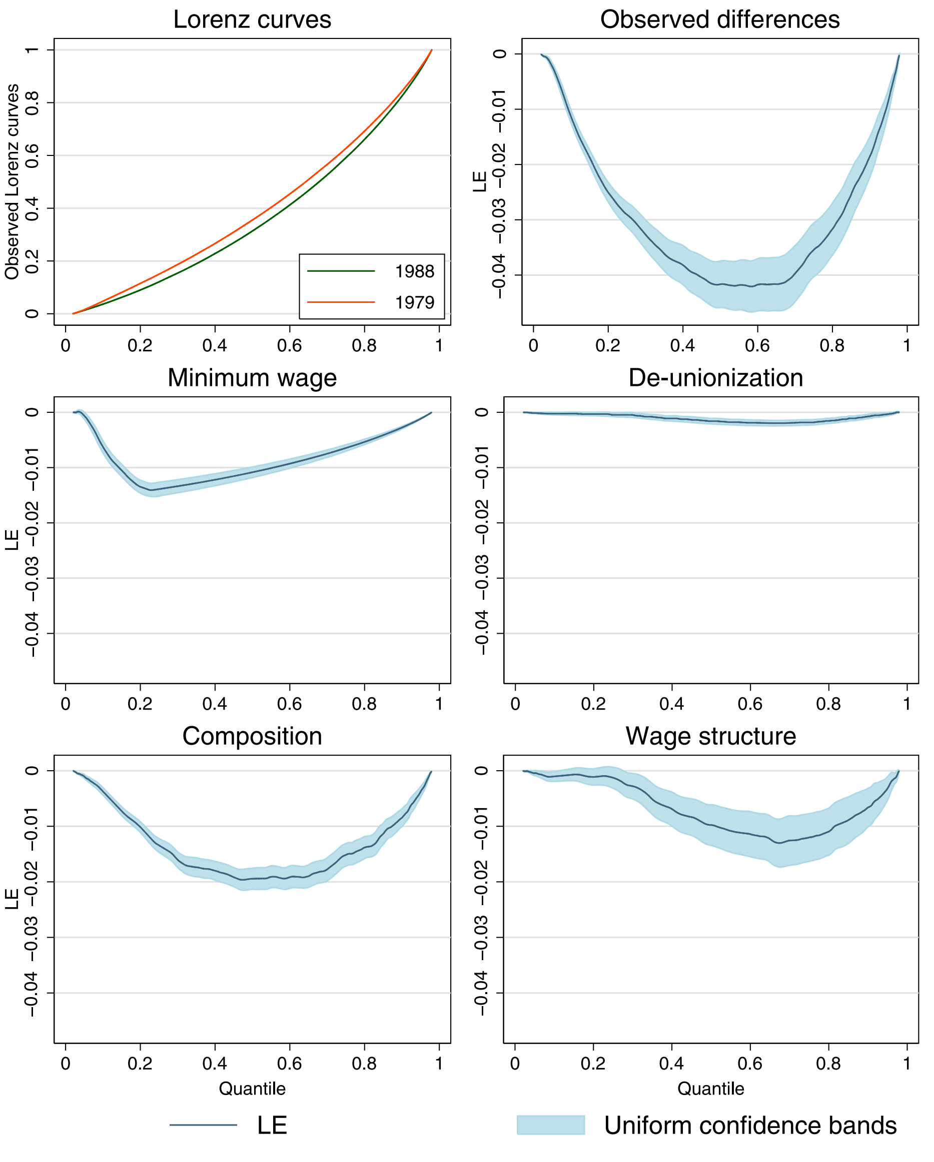

FIGURE SC.5.—Observed Lorenz curves, observed differences between the Lorenz curves and their decomposition into four Lorenz effects. The 95% simultaneous confidence bands were obtained by empirical bootstrap with 100 repetitions. Results for women.

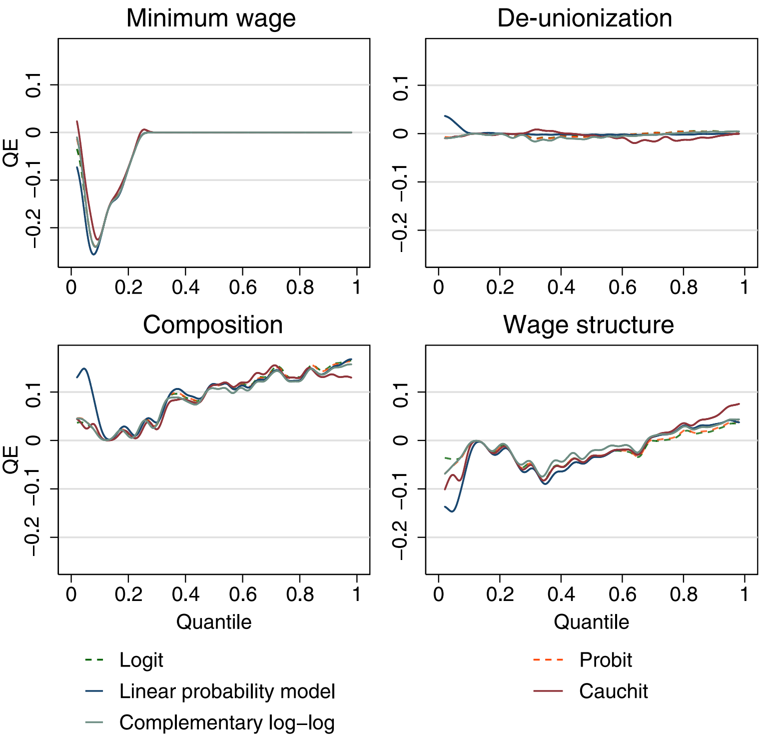

FIGURE SC.6.—Comparison of the distribution regression estimates of the quantile effects based on five different link functions: logistic, normal, uniform, Cauchy and complementary log-log. Results for women.

where the first term is the between-group variance and the second term is the within-group variance. The regression models for conditional distributions considered in CFM naturally lead to models for the conditional mean and variance of Y given X. By definition,

$$E(Y|X=x)=\int y\,dF_{Y|X}(y|x)\quad\text{and}$$

$$\mathrm{Var}(Y|X=x)=\int\big(y-E(Y|X=x)\big)^2\,dF_{Y|X}(y|x).$$

For conditional quantile models, the direct expressions are

$$E(Y|X=x)=\int_0^1 Q_{Y|X}(u|x)\,du\quad\text{and}$$

$$\mathrm{Var}(Y|X=x)=\int_0^1\big(Q_{Y|X}(u|x)-E(Y|X=x)\big)^2\,du.$$

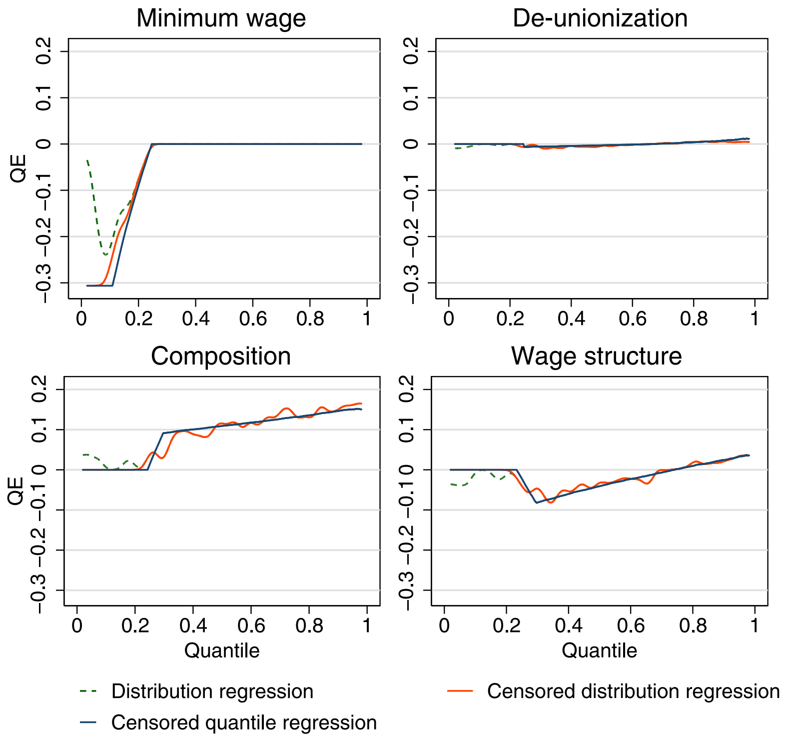

FIGURE SC.7.—Comparison of the distribution regression, censored distribution regression and censored quantile regression estimates of the quantile effects. Results for women.

For linear quantile regression models, this simplifies to

$$E(Y|X=x) = x'E\big[\beta(u)\big] \quad \text{and}$$

$$\operatorname{Var}(Y|X=x) = x' \operatorname{Var}\big[\beta(u)\big]x,$$

such that the variance decomposition takes the simple form (equation (6.7) in CFM)

$$\operatorname{Var}[Y] = E\big[\beta(U)\big]' \operatorname{Var}[X]E\big[\beta(U)\big] + \operatorname{trace}\big\{E\big[XX'\big]\operatorname{Var}\big[\beta(U)\big]\big\}.$$

Both the between and the within components are functions of the conditional distribution $F_{Y|X}$ (that determines $E(Y|X=x)$ and $\operatorname{Var}(Y|X=x)$) and the covariate distribution F_X. Our counterfactual changes consists of changing F_X while keeping $F_{Y|X}$ fixed. In general, the between and within components of the variance of the counterfactual outcome $Y\langle j, k\rangle$ are

$$\operatorname{Var}\big[E\big(Y\langle j, k\rangle|X\big)\big] = \int \big[E(Y_j|X_j = x) - E\big(Y\langle j, k\rangle\big)\big]^2 \, dF_{X_k}(x) \quad \text{and}$$

$$E\big[\operatorname{Var}\big(Y\langle j, k\rangle|X\big)\big] = \int \operatorname{Var}(Y_j|X_j = x)\, dF_{X_k}(x).$$

In Section 6 of CFM, the whole composition effect (i.e., the effect of changing the distribution of both union status and other individual characteristics) is defined as the difference between $F_{Y\langle(1,0)|(1,1)\rangle}$ and $F_{Y\langle(1,0)|(0,0)\rangle}$. Therefore, the composition effects on between-group and within-group inequality are

$$\operatorname{Var}\big[E\big(Y\big\langle(1,0)|(1,1)\big\rangle\big|X\big)\big] - \operatorname{Var}\big[E\big(Y\big\langle(1,0)|(0,0)\big\rangle\big|X\big)\big] \quad \text{and}$$

$$E\big[\operatorname{Var}\big(Y\big\langle(1,0)|(1,1)\big\rangle\big|X\big)\big] - E\big[\operatorname{Var}\big(Y\big\langle(1,0)|(0,0)\big\rangle\big|X\big)\big].$$

When we use the logit distribution regression model to estimate the conditional distribution $F_{Y_{(1,0)}|X_1}$, we find that between-group inequality increased from 0.147 to 0.163 as a consequence of the composition changes, which represents an increase of 10.6 percent. Similarly, within-group inequality increased from 0.125 to 0.136, which represents an increase of 9.1 percent. Note that the dependent variable is the hourly log wage.

When we use the censored quantile regression model to estimate the conditional distribution that censors wages below the minimum wage, we obtain increases in between-group and within-group inequality from 0.135 to 0.145 (7.5 percent) and from 0.120 to 0.131 (9.1 percent), respectively. If we do not censor wages below the minimum wage, then we can use directly the expressions in equation (6.7). In this case, we obtain increases for between-group and within-group inequality from 0.201 to 0.217 (7.9 percent) and from 0.152 to 0.164 (8.4 percent), respectively.

REFERENCES

BERAN, R. (1984): "Bootstrap Methods in Statistics," *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 86, 14–30. [4]

CHERNOZHUKOV, V., AND I. FERNÁNDEZ-VAL (2005): "Subsampling Inference on Quantile Regression Processes," *Sankhyā*, 67, 253–276. [2,3]

CHERNOZHUKOV, V., AND H. HONG (2002): "Three-Step Censored Quantile Regression and Extramarital Affairs," *Journal of the American Statistical Association*, 97, 872–882. [15]

CHERNOZHUKOV, V., I. FERNÁNDEZ-VAL, AND A. GALICHON (2010): "Quantile and Probability Curves Without Crossing," *Econometrica*, 78 (3), 559–575. [7]

DAVYDOV, Y. A., M. A. LIFSHITS, AND N. V. SMORODINA (1998): *Local Properties of Distributions of Stochastic Functionals*. Translations of Mathematical Monographs, Vol. 173. Providence, RI: American Mathematical Society. Translated from the 1995 Russian original by V. E. Nazaĭkinskiĭ and M. A. Shishkova. [4]

KOENKER, R. (2010): "On Distributional vs. Quantile Regressions," Unpublished Manuscript, University of Illinois at Urbana–Champaign. [12]

KOENKER, R., AND G. BASSETT (1978): "Regression Quantiles," *Econometrica*, 46, 33–50. [7]

POWELL, J. L. (1986): "Censored Regression Quantiles," *Journal of Econometrics*, 32, 143–155. [15]

ROTHE, C., AND D. WIED (2012): "Misspecification Testing in a Class of Conditional Distributional Models," *Journal of the American Statistical Association*, 108, 314–324. [12]

VAN DER VAART, A., AND J. WELLNER (1996): *Weak Convergence and Empirical Processes: With Applications to Statistics*. New York: Springer. [2]

Dept. of Economics, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.; vchern@mit.edu; http://www.mit.edu/~vchern/,

Dept. of Economics, Boston University, Boston, MA 02215, USA; ivanf@bu.edu; http://people.bu.edu/ivanf/,

and

Dept. of Economics, Bern University, 3001 Bern, Switzerland; blaise.melly@vwi.unibe.ch; http://staff.vwi.unibe.ch/melly/.

Manuscript received April, 2009; final revision received November, 2012.